R E P O R T R E S U M E S

ED 020 082 RE 001 240

READING TESTS AND TEACHERS.

BY- FARR, ROGER

PUB DATE APR 68

EDRS PRICE MF-$0.25 HC-$0.44 9P.

DESCRIPTORS- *READING TESTS, *MEASUREMENT GOALS, INFORMAL READING INVENTORY, READING DIAGNOSIS, READING SKILLS, READING ACHIEVEMENT, OBJECTIVES, EVALUATION TECHNIQUES, TESTING, *EVALUATION,

TESTS SHOULD BE CAREFULLY EXAMINED IF TEACHING OBJECTIVES ARE TO MATCH TEST OBJECTIVES. TEACHERS SHOULD BECOME MORE PROFICIENT IN THE BROADER ASPECTS OF EVALUATING READING PERFORMANCE AND SHOULD EMPLOY A WIDER VARIETY OF EVALUATION TECHNIQUES. THE USE OF STANDARDIZED READING TESTS AS INDICATORS OF INSTRUCTIONAL READING LEVEL SHOULD BE ABANDONED. PUBLISHERS SHOULD BE FORCED BY TEST USERS TO DISCOVER THAT IT IS UNPROFITABLE TO PUBLISH A TEST THAT DOES NOT MEET THE AMERICAN PSYCHOLOGICAL ASSOCIATION'S MINIMUM STANDARDS. SUBTEST SCORES OF STANDARDIZED READING TESTS SHOULD BE INTERPRETED WITH CONSIDERABLE CAUTION. EIGHT REFERENCES ARE GIVEN. THIS PAPER WAS PRESENTED AT THE INTERNATIONAL READING ASSOCIATION CONFERENCE (BOSTON, APRIL 24-27, 1968). (AUTHOR/RJ)

[INTERNATIONAL READING ASSOCIATION, BOSTON, APRIL 24-27, 1968]

Dr. Roger Farr
Assistant Professor of Education
222 Pine Hall
Indiana University
Bloomington, Indiana 47401

ED020082

Reading Tests and Teachers

Session 11A--Measurement in Reading

It does not take an extremely perceptive observer of classroom teachers to discover that a large percentage of teachers are quite naive in regard to even the simplest and most basic principles of evaluation. This factor is perhaps the single most important impediment to the existence of better reading tests. Oscar Buros in the Preface to the Sixth Mental Measurements Yearbook published in 1965 makes this point quite cogently:

> Unfortunately, the rank and file of test users do not appear to be particularly alarmed that so many tests are either severely criticized or described as having no validity. Although most test users would probably agree that many tests are either worthless or misused, they continue to have the utmost faith in their own particular choice and use of tests regardless of the absence of supporting research or even of the presence of negating research. When I initiated this test

RE001 240

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

> reviewing service in 1938, I was confident that frankly critical reviews by competent specialists representing a wide variety of viewpoints would make it unprofitable to publish tests of unknown or questionable validity. Now 27 years and five Mental Measurement Yearbooks later, I realize that I was too optimistic. (2)

Buros then goes on to say that he is more confident than ever of the correctness of the following statement which was written for the Introduction to Tests in Print:

> At present, no matter how poor a test may be, if it is nicely packaged and if it promises to do all sorts of things which no test can do, the test will find many gullible buyers. (2)

This author strongly suggests that poor reading tests of the past were produced and published because test consumers did not demand higher quality from test publishers or developers. Furthermore, there is little evidence to suggest that the situation is any better today or that it will be markedly improved in the future.

In order to determine whether reading tests are being used effectively it is important to examine the reasons that tests are administered. The primary purpose is to provide the instructor with feedback information regarding student progress toward goals. In regard to this point it is also important that students should be kept informed of their progress. A second major reason for administering tests is to evaluate student strengths and weaknesses so an effective instructional plan can be developed. There are other reasons, however, they are of minor significance by comparison. Tests might be used as part of a research study, to justify the existence of a program to an administrator or to determine retention of gains or transfer of reading skills to other academic areas.

What problems, inherent in reading tests and the use of these tests by teachers, mitigate against the use of tests for providing feedback information regarding student progress toward goals? First of all, a teacher should list the objectives of his instructional program and then secure several tests which seem to fit his purposes. The next step would be to ferret out the objectives of each test. This can be done by critically examining the author's statements about the test, the information in the test manual, and each sub-test and individual item on the test. Critical reviews of the tests from Buros' Mental Measurement Yearbooks (2) can also aid in this task. Once the objectives of each test have been determined, the selection of the best test can be made on the basis of how closely the teaching objectives and the test objectives compare.

A word of caution should be raised regarding the usual cursory review of test items made by most test consumers. Merely because a test has a vocabulary sub-test and one of the teaching objectives is the development of students' vocabularies, it does not follow that the test meets this particular teaching objective. Other points to be considered include whether or not the vocabulary test is timed, whether vocabulary words are presented in context or in isolation, whether the subject matter is specific or general, and whether spelling of homonyms is included among the discriminations to be made. In regard to each of these points, the teacher needs to determine whether or not this is indeed the objective he has attempted to develop with students.

If teachers go through the process of matching teaching objectives to test objectives, they will quickly come to several conclusions. The

first is that the method by which a particular teacher attempts to develop a specific reading skill and the way in which that skill is measured on a specific reading test are probably quite different. Secondly, teachers will become quite cognizant of the fact that many of the objectives which they have indicated as being quite important are not measured on any standardized reading test. These include improving student attitudes toward reading, increasing the habit of reading, broadening interests in reading and applying study skills.

The problem of lack of tests for several important objectives should cause teachers to begin planning strategies by which progress toward objectives can be measured. These strategies cannot, of course, rely on standardized tests: skills check sheets, anecdotal records, progress in workbooks, and other techniques can be refined to give a valid and reliable description of student progress toward these important objectives. A collection of interesting techniques for informal measurement are described in a book entitled Unobtrusive Measures (8).

Finally the definition of progress toward an objective must be faced. Once a test has been administered or a skill has been measured in some informal manner, the teacher needs to decide if a particular performance represents progress. Fred B. Davis has outlined four steps that he feels are necessary in measuring change:

1. It is first necessary to define carefully and explicitly the variable, being measured.
2. Second, a test of the variable, or as close an approximation to it as can be secured, must be administered under conditions that assure a high degree of cooperation on the part of the pupil.

3. Third, a pupil's obtained score must be compared with suitable norms, such as percentile ranks in his own age or grade group.
4. Fourth, the possibility that the pupils obtained score represents a sizeable deviation from his true score must be considered. (3)

An additional point should be added whenever we are concerning ourselves with evaluating progress and are primarily using test scores. That is that a student's score on any reading skill test represents one sampling of behavior, on one operational definition of that skill, under one specific set of conditions, at one particular point in a students development and that that is really needed for valid and reliable evaluation is the sampling of many behaviors, on many different aspects of the skill, under many sets of conditions, at many points in a student's development.

As stated previously the second major reason that tests are administered is to determine student's strengths and weaknesses so an effective instructional program can be planned. How well do reading tests and teacher use of them aid in this objective?

First of all there is almost no evidence available that any of the sub-tests of standardized reading tests are valid measures of separate and distinct reading skills. This fact should cast a great deal of concern on the common use of standardized reading tests as diagnostic tools. If a particular test does not validly measure distinct sub-skills of reading, certainly an instructional program based on an analysis of students' sub-test performance would be subject to serious question. This problem, however, does not hamper

the sales of tests. By examining the popularity of various tests it is safe to conclude that those tests which provide more sub-test scores and promise to be more diagnostic are the best sellers.

There have been a plethora of investigations dealing with the validity of reading tests. Most of these have attempted to isolate specific factors of reading behavior, often employing the factor analytic technique. Lennon reviewed twelve of these investigations in a 1962 article entitled, "What Can Be Measured?" (6) In concluding his article Lennon stated: "It would seem that we may recognize and hope to measure reliably the following components of reading ability: (1.) a general verbal factor, (2.) comprehension of explicitly stated material, (3.) comprehension of implicit or latent meaning, and (4.) an element that might be termed 'appreciation'."

A second misuse of reading tests in the diagnosis of students' reading ability is the use of the grade level score as an indication of the level at which instruction should be provided. There have been a number of studies (5) (7) which have consistently demonstrated that a student's score on a standardized reading test and his performance on an informal reading inventory are not comparable. These studies have found that most often the standardized test score is a more consistent indicator of the student's frustration reading level than his instructional reading level. For a discussion of these functional reading levels, the reader is referred to Bett's classic description in Foundations of Reading Instruction (1).

There are a number of ways that standardized reading tests can help the teacher to plan his instructional program. First of all, despite the fact that the standardized reading tests do not place a youngster at his appropriate instructional level there is ample evidence that the ranking of students' performance on standardized reading tests as compared to informal reading tests does not differ significantly. The standardized test could therefore be used to place youngsters in general ability level reading groups and could also be used as a screening device for students who are in need of a more extensive evaluation.

Teachers could also learn more about their students if they would examine each student's individual responses. Many times teachers are only given the sub-test and total test scores for each student and never have an opportunity to match student responses with the questions on the test. A careful matching of questions and answers can often reveal a great deal of diagnostic information.

The evaluation of student's reading performance can be greatly improved if test consumers would consider the following point:

1. A careful examination of the test should be made to match teaching objectives to test objectives.

2. Teachers should become more proficient in the broader aspects of evaluating reading performance. A wider variety of evaluation techniques should be employed.

3. The use of standardized reading tests as indicators of instructional reading level should be abandoned.

4. Publishers should be forced by test consumers to discover that it is improfitable to publish a test that does not meet the APA (4) minimum standards.

5. Sub-test scores of standardized reading tests should be interpreted with considerable caution.

1. Betts, Emmett Albert. Foundations of Reading Instruction. New York: American Book Company, 1957.

2. Buros, Oscar K. The Sixth Mental Measurements Yearbook. Highland Park, New Jersey: The Cryphon Press, 1965.

3. Davis, Fred B. "The Assessment of Change," Phases of College and Other Adult Reading Programs: Tenth Yearbook of the National Reading Conference. Milwaukee, Wisconsin: The National Reading Conference, 1961.

4. French, John W. and Michael, William B., Co-chairmen. Standards for Educational and Psychological Tests and Manuals. Washington, D.C.: American Psychological Association, Inc., 1966.

5. Glasser, N.A. A Comparison of Specific Reading Skills of Advanced and Retarded Readers of Fifth Grade Reading Achievement. Unpublished doctoral dissertation, University of Oregon, 1964.

6. Lennon, Roger. "What Can Be Measured?" The Reading Teacher, 15 (March, 1962), 326-37.

7. Sipay, E. R. "A Comparison of Standardized Reading Scores and Functional Reading Levels," The Reading Teacher, 17 (January, 1964), 265-68.

8. Webb, Eugene J. et. al. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally and Company, 1966.